Faculty Working Papers

ON DUMMY VARIABLES

Lucy Chao Lee

#113

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPER

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

May 24, 1973

ON DUMMY VARIABLES

Lucy Chao Lee

#113

On Dummy Variables

Lucy Chao Lee

Statement of Problem

It is well known that when sets of dummy variables are included in a regression function, multicollinearity with the constant term prevents the least-squares estimation of the coefficients. The common practice to circumvent this difficulty is to exclude any one of the dummy variable in a set, and then proceed merrily with the estimation. However, one might be interested in comparing the relative importance of all the dummy variables in a set for the prediction. Therefore, the question arises as to the estimation of the coefficient of the excluded dummy variable and the different effect of excluding one or the other dummy variable in the set.

The objective of this paper is to review the known facts on using one set of dummy variables in regression analysis, and investigate the advantage of applying the principal components technique to the general case of more than one set of dummy variables. Two empirical examples will be used to illustrate the behavior of least-squares estimates of the coefficients of dummy variables for various specifications of a function.

The Simple Case

In the simple case of only one set of dummy variables with other independent variables in the equation, these problems have been analyzed.[1] The approach used is to exclude the constant term from the original specification to allow estimation of the coefficients of all the dummy variables

---

[1] Goldberger, Arthur S. Econometric Theory (New York: John Wiley and Sons, Inc., 1964), pp. 218-227.

in the set. Then any one dummy variable is omitted in the second specification, and the constant term included. A comparison of the least-squares estimates of the coefficients of the regressors in the second specification with those in the original reveals the following facts, the proofs of which are given in Appendix I for completeness.

1. The constant term is the coefficient of the dummy variable in the original specification now omitted.

2. The coefficients of the included dummy variables in the second specification are the differences of their original coefficients and that of the omitted dummy variable.

3. The coefficients of other non-dummy regressors, if any, remain unchanged.

4. Both estimated functions will give the same estimated values of the regressand and therefore R^2 remains the same.

General Case and Application of Principal Components Analysis

If more than one set of dummy variables is included in the equation, multicollinearity exists even without using the constant term. Any non-square transformation may be used to exclude any one dummy variable from each set and include the constant term. As a result, one cannot find a unique specification as in the prior simple case, as a basis for finding the functional relationship between the estimated coefficients in all the possible transformed equations. Even though all these estimated functions will give the same estimated dependent variable and the same R^2, question arises as to the relative importance of all the dummy variables in a set. This problem, well known in quantitative psychology, can be approached by principal components analysis.

Principal components analysis is a technique to find a smaller set of variables, the principal components, in a linear function to explain each of the original set of variables. Especially where near-degeneracy exists in the original data, replacing them with the principal components results in condensation of information. Finding the principal components may be an end in itself, and it may be used as a first-stage solution to factor analysis or a preliminary to regression analysis[2] which is the case in this study.

Since multicollinearity exists among sets of dummy variables and the constant term, principal components technique enables us to extract from them a smaller set of independent variables that reproduce all the data variation in them. This new set of independent variables, the principal components, can then be used as regressors instead of the original for explaining the dependent variable. It is shown in Appendix II that the coefficients or all the original regressors can be obtained by a linear transformation of the estimated coefficients of the principal components. Thus, an assessment of the relative importance of all the original regressors is permitted. Furthermore, for prediction it is not necessary to convert the original regressors into principal components.

Examples

The prior theoretical material can be best illustrated by an example. The data are based on a subsample of 38 families that purchased automobiles

---

[2]Tatsuoka, Maurice M. Multivariate Analysis (New York: John Wiley and Sons, Inc., 1971), pp. 144-149.

in 1969, part of a panel of young couples maintained by SRL in Peoria and Decatur, Illinois. The selected variables include the dependent variable y, independent variables x_1, x_2, and a set of dummy variables d_1, d_2, and d_3, where

y = price of automobile in hundreds of dollars

x_1 = husband's education in number of years

x_2 = 1969 family income in thousands of dollars

d_1 = husband assumes the role of financial officer

d_2 = wife assumes the role of financial officer

d_3 = husband and wife jointly assume the role of financial officer

Single equation least-squares is used to estimate the coefficients in various specifications of the function to explain the price paid for an automobile. The regression results are summarized in Table 1, each column depicting a separate function.

Note that R^2 and the estimated coefficients of x_1 and x_2 are the same in functions (3) through (7). In function (7) the coefficients of all the regressors are obtained by a linear transformation of the estimated coefficients of the principal components.

It is evident that the constant terms in function (4), (5), and (6) are respectively the estimated coefficients of the dummy variables in (3) but omitted in (4), (5) and (6). Furthermore, the coefficients of the two included dummy variables respectively in (4), (5), and (6) are the differences between the estimated coefficients of the corresponding dummy variables in (3) and the constant terms in (4), (5) and (6).

Table 1 Estimated Function of Automobile Price with One Set of Dummy Variables

Variable	Regression									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant		-14.93		-17.27	-14.32	-18.01	-12.39	-13.25	-18.07	-15.96
		(-0.85)		(-0.89)	(-0.79)	(-0.98)	(-0.90)	(-0.74)	(-1.00)	(-0.89)
x_1	0.29	1.38	1.47	1.47	1.47	1.47	1.47	1.20	1.53	1.58
	(0.53)	(0.99)	(0.95)	(0.96)	(0.97)	(0.97)	(0.96)	(0.83)	(1.09)	(1.05)
x_2	1.22	1.31	1.33	1.33	1.33	1.33	1.33	1.43	1.30	1.23
	(2.02)	(2.16)	(1.94)	(1.97)	(1.97)	(1.97)	(1.97)	(2.20)	(2.12)	(1.88)
d_1			-18.01	-0.74	-3.68		-5.60	-2.66		
			(-0.97)	(-0.13)	(-0.71)		(-1.05)	(-0.55)		
d_2			-14.33	2.94		3.68	-1.93		3.29	
			(-0.78)	(0.59)		(0.71)	(-0.38)		(0.81)	
d_3			-17.27		-2.94	0.74	-4.87			-1.75
			(-0.88)		(-0.59)	(0.13)	(0.79)			(-0.38)
R^2	0.126	0.162	0.178	0.178	0.178	0.178	0.178	0.170	0.178	0.166
RSS	769.129	990.586	1089.391	1089.391	1089.391	1089.391	1089.391	1036.177	1086.947	1011.983
ESS	5340.404	5118.947	5020.142	5020.142	5020.142	5020.142	5020.142	5073.356	5022.586	5097.550
$S^2 = \frac{ESS}{df.}$	148.345	146.241	152.128	152.128	152.128	152.128	152.128	149.206	146.720	149.916

Note: Figures in parentheses are t values of coefficients just above.

Comparing the RSS (regression sum of squares) of function (1) with those of (3), (4), (5) and (6), it is noted that the additional contribution of all three dummy variables in (3) is the same as that of any two dummy variables and the constant term in (4), (5) and (6). Similarly, comparing the RSS of function (2) with those of (4), (5) and (6), it is also evident that the additional contribution of any two dummy variables in (4), (5) and (6) is the same.

It is of interest to observe that the coefficients of d_3 in functions (5) and (6) bear different signs. Therefore, in interpreting the regression results of any of the functions (4), (5) and (6), the researcher should be aware that the positive or negative contribution of an included dummy variable is only relative to the omitted dummy variable in the set.

In empirical studies, it is a common practice to select only variables with higher t-ratios in the initial regression results for a second regression fit, with a view to gaining more degrees of freedom with little sacrifice in goodness of fit. However, one should be cautious in applying this rule to dummy variables. A dummy variable dropped from the function is actually combined with the dummy variable initially excluded from the function, thereby forming a smaller set of dummy variables. This new set of dummy variables may or may not be the optimal set among all the alternative sets for the second fit in terms of goodness of fit and mean squared error, since the initial function is not unique.

For example, suppose out of alternatives (4), (5) and (6), function (5) was chosen to give the initial regression results, and dummy variable d_3 with a t-ratio of -0.59 is dropped for the second run in (8). The results of (8) show that R^2 and s^2 (mean squared error) are not as good as

those of (9) in which dummy variable d_1 is dropped. On the other hand, suppose function (4) was chosen to give the initial results, the t-ratio rule would lead to omission of d_1 and the optimal second fit in (9). In other words, an included dummy variable in the initial function with a higher t-ratio may have a lower t-ratio when a different dummy variable is excluded from an alternative initial function. For example, compare the t-ratios of d_1 in (4) and (5).

Therefore, some other criterion is needed for combining the dummy variables in order to gain more degrees of freedom and reduce mean squared error in the second fit. It seems the results given by the principal components approach in (7) might throw light upon achieving the purpose. Since the coefficients of all the dummy variables are obtained, comparison among them may be helpful. It is observed that the coefficients of d_1 and d_3 in (7) has the smallest difference among all possible three pairs of coefficients in the set. It is felt that combining these two dummy variables would constitute the set that can best serve the purpose. This hypothesis is borne out by the results in (9). However, no mathematical proof is attempted in this paper.

To illustrate the general case of two or more sets of dummy variables, a set of two dummy variables e_1 and e_2 is added to the previous model of automobile price, where

e_1 = wife works

e_2 = wife does not work

Again, least-squares estimates are run for various specifications of the function, and are displayed in Table 2.

Table 2 Estimated Function of Automobile Price with Two Sets of Dummy Variables

Variable	Regression											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant						-18.80	-14.20	-16.65	-10.65	-12.65	-16.64	-15.24
						(-0.98)	(0.79)	(-0.92)	(-1.06)	(-0.71)	(-0.93)	(-0.86)
x_1	1.91	1.91	1.91	1.91	1.91	1.91	1.91	1.91	1.91	1.43	1.70	2.01
	(1.22)	(1.22)	(1.22)	(1.22)	(1.22)	(1.23)	(1.23)	(1.23)	(1.23)	(0.98)	(1.21)	(1.33)
x_2	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.22	1.38	1.32	1.15
	(1.78)	(1.78)	(1.78)	(1.78)	(1.78)	(1.81)	(1.81)	(1.81)	(1.81)	(2.13)	(2.17)	(1.77)
d_1	-16.65	-22.62	-2.15	-2.46		2.15	-2.46		-3.66	-1.19		
	(-0.90)	(-1.20)	(-0.34)	(-0.47)		(0.34)	(-0.47)		(-0.83)	(-0.24)		
d_2	-14.20	-20.16	4.60		2.46	4.60		2.46	-1.21		3.56	
	(-0.77)	(-1.07)	(0.89)		(0.47)	(0.90)		(0.47)	(-0.30)		(0.88)	
d_3	-18.80	-24.77		-4.60	-2.15		-4.60	-2.15	-5.81			-3.95
	(-0.96)	(-1.21)		(-0.89)	(-0.34)		(-0.90)	(-0.34)	(-1.09)			(-0.81)
e_1	-5.97		-24.77	-20.16	-22.62	-5.97	-5.97	-5.97	-8.31	-4.87	-5.37	-6.39
	(-1.24)		(-1.21)	(-1.07)	(-1.20)	(-1.26)	(-1.26)	(-1.26)	(-1.43)	(-1.07)	(-1.24)	(-1.39)
e_2		5.97	-18.80	-14.20	-16.65				-2.34			
		(1.24)	(-0.96)	(-0.77)	(-0.90)				(-0.46)			
R^2	0.217	0.217	0.217	0.217	0.217	0.217	0.217	0.217	0.217	0.197	0.214	0.212
RSS	1327.296	1327.296	1327.296	1327.296	1327.296	1327.296	1327.296	1327.296	1327.296	1205.710	1309.581	1294.014
ESS	4782.237	4782.237	4782.237	4782.237	4782.237	4782.237	4782.237	4782.237	4782.237	4903.823	4799.952	4815.519
$S^2 = \frac{ESS}{df.}$	149.426	149.426	149.426	149.426	149.426	149.426	149.426	149.426	149.426	148.596	145.444	145.926

Note that all nine of these estimated functions in the first nine columns yield the same R^2 as well as the same coefficients of x_1 and x_2. In function (9), the principal components approach is applied to obtain the coefficients of all the regressors.

When the constant term is omitted in the function, only one dummy variable has to be excluded from either set. The results of the five possible choices are shown in columns (1) through (5). It is interesting to note that the estimated coefficients of e_1 and e_2 in (3) are those of d_3 respectively in (1) and (2). When the constant term is specified in the function, one dummy variable has to be excluded from each set. Only three of the possible six functions are shown in column (6), (7) and (8). Note that the constant term in (6) is the estimated coefficient of d_3 in (1), and that the estimated coefficients of d_1 and d_2 in (6) are respectively the differences between those in (1) and the constant term in (6).

Similar to the case of using only one set of dummy variables, the results of (9) can throw light upon the choice of combining dummy variables in a set into a new smaller set. Note that the difference between the estimated coefficients of d_1 and d_3 in (9) is the smallest among the possible three pairs. The results of (11) support the theory that combining d_1 and d_3 yields the best R^2 among all the three possible choices of combination.

Summary

This paper is an attempt to synthesize some known facts about dummy variables in regression analysis. Some examples are used to display the behavior of the estimated coefficients of the dummy variables in alternative

specifications. It is pointed out that using the usual criterion of t-ratios to delete dummy variables may not bring about the best goodness of fit unless the initial specification is the lucky one among the alternatives. In addition, the well-known technique of principal components analysis is applied to the problem, to circumvent the non-uniqueness of specification in using sets of dummy variables. It is conjectured that the coefficients thus obtained for all the dummy variables may be useful for the best choice of combining dummy variables in terms of goodness of fit.

Appendix I

Assume that there are N observations in k independent variables $x_1, x_2, \ldots, x_k$ and without loss of generality, a set of three dummy variables d_1, d_2, and d_3. To avoid multicollinearity and form a unique basis for later comparisons, the constant is excluded from the initial specification. Using matrix notation, the N equations are written

$$Y = X\beta + \mu$$

$$= [D_1, D_2, D_3, X_k] \begin{bmatrix} \beta_{d_1} \\ \beta_{d_2} \\ \beta_{d_3} \\ \beta_x \end{bmatrix} + \mu$$

where X is the N x (k+3) total regressor matrix of N observations on $d_1, d_2, d_3, x_1, x_2, \ldots, x_k$

Y is the N x 1 vector of observations on the regressand

D_1, D_2 and D_3 are respectively the N x 1 vector of observations on d_1, d_2 and d_3

X_k is the N x k matrix of observations on the k non-dummy regressors $x_1, x_2, \ldots x_k$

β_{d_1}, β_{d_2} and β_{d_3} are respectively the coefficients of d_1, d_2 and d_3,

β_x is the k x 1 vector of coefficients of the x's, and

μ is the N x 1 vector of disturbances, with $E(\mu) = 0$ and $\text{Var}(\mu) = \sigma^2 I$.

The least-squares estimate of the coefficients is given by

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_{d_1} \\ \hat{\beta}_{d_2} \\ \hat{\beta}_{d_3} \\ \hat{\beta}_x \end{bmatrix} = (X\,X)^{-1}\,X\,Y$$

Next, let the specification be changed by arbitrarily omitting dummy variable d_1 and substituting a constant term c. The new regressor matrix can easily be obtained by applying a (k+3) x (k+3) non-singular square transformation matrix to the original total regressor matrix X, i.e.

$$Z = XT$$

where

$$T = \left[\begin{array}{ccc:c} 1 & 0 & 0 & \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & \\ \hdashline & 0 & & I_k \end{array}\right]$$

and where I_k is a k x k identity matrix. In this new regression function $Y = Z\beta + \mu$, the least-squares estimate of the coefficients for the transformed regressor Z is

$$\begin{aligned} \tilde{\beta} &= (Z'Z)^{-1} Z'Y \\ &= (T'X'XT)^{-1} T'X'Y \\ &= T^{-1} (X'X)^{-1} T'^{-1}T'X'Y \\ &= T^{-1} (X'X)^{-1} X'Y \\ &= T^{-1} \hat{\beta} \end{aligned}$$

Since the inverse of T is found as

$$T^{-1} = \left[\begin{array}{ccc:c} 1 & 0 & 0 & \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & \\ \hdashline & 0 & & I_k \end{array}\right]$$

substituting T^{-1} and $\hat{\beta}$ into the equation for $\tilde{\beta}$ gives the relationship between the components of $\tilde{\beta}$ and $\hat{\beta}$

$$\tilde{\beta} = \begin{bmatrix} \tilde{\beta}_c \\ \tilde{\beta}_{d_2} \\ \tilde{\beta}_{d_3} \\ \tilde{\beta}_x \end{bmatrix} = \left[\begin{array}{ccc:c} 1 & 0 & 0 & \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & \\ \hdashline & 0 & & I_k \end{array}\right] \begin{bmatrix} \hat{\beta}_{d_1} \\ \hat{\beta}_{d_2} \\ \hat{\beta}_{d_3} \\ \hat{\beta}_x \end{bmatrix} = \begin{bmatrix} \hat{\beta}_{d_1} \\ \hat{\beta}_{d_2} - \hat{\beta}_{d_1} \\ \hat{\beta}_{d_3} - \hat{\beta}_{d_1} \\ \hat{\beta}_x \end{bmatrix}$$

Finally, the estimated values of the regressand will be affected by the change in specification as shown

$$\tilde{Y} = Z\tilde{\beta} = XTT^{-1}\hat{\beta} = X\hat{\beta} = \hat{Y}$$

Appendix II

Let X be the raw data regressor matrix of N observations on (k + p + 1) regressors, i.e., the constant term, k non-dummy independent variables, and p dummy variables in m sets. Since there is linear dependency among the dummy variables and the constant term, the rank of X is (k + p + 1 - m). Therefore, it is possible to find a set of (k + p + 1 - m) variables, smaller than the original set of (k + p + 1) variables, that reproduce all the data variation in X. This new set of variables Z, or principal components, can then be used as regressors in place of X for explaining Y.

Using principal components analysis, one can find a transformation V_* such that $Z = XV_*$, by diagonizing $X'X$ which is a (k + p + 1) x (k + k + 1) symmetric matrix. That is,

$$X'X = V\Lambda V'$$

where Λ is a (k + p + 1) x (k + p + 1) diagonal matrix of eigen values of $X'X$

V is a (k + p + 1) x (k + p + 1) matrix of corresponding eigen vectors and $V'V = I$

Since the rank of $X'X$ is (k + p + 1 - m), one can only find (k + p + 1 - m) positive eigen values in Λ, i.e., the last m eigen values on the diagonal of Λ are zeroes. Thus, a set linearly independent variables Z can be found by using the first (k + p + 1 - m) eigen vectors in V as a transformation V_*. The principal components are

$$Z = XV_*$$

where Z is N x (k + p + 1 - m)

X is N x (k + p + 1)

V_* is (k + p + 1) x (k + p + 1 - m)

One can then estimate the parameters for the principal components in the following model

$$Y = Z\beta_z + \mu$$

where Y is the $N \times 1$ vector of regressand

Z is the $N \times (k + p + 1 - m)$ matrix of derived principal components

β_z is the $(k + p + 1 - m) \times 1$ vector of coefficients of the principal components, and

μ is the $N \times 1$ vector of disturbances with $E(\mu) = 0$ and $Var(\mu) = \sigma^2 I$.

Applying the least squares method to the above model gives the following two results:

1. The estimate of β_z is

$$\hat{\beta}_z = (Z'Z)^{-1} Z'Y = (V_*' X'X V_*)^{-1} V_*'X'Y$$
$$= (V_*'V\Lambda V'V_*)^{-1} V_*'X'Y$$
$$= \Lambda_*^{-1} V_*'X'Y$$

where Λ is the $(k + p + 1 - m) \times (k + p + 1 - m)$ diagonal matrix of the first $(k + p + 1 - m)$ positive eigen values in Λ.

2. The estimated values of the regressand are

$$\hat{Y} = Z \hat{\beta}_z = (XV_*) \hat{\beta}_z$$
$$= X(V_*\hat{\beta}_z)$$
$$= X\hat{\beta}_x$$

where $\hat{\beta}_x$ is defined to be $V_*\hat{\beta}_z$ or $V_*(\Lambda_*^{-1}V_*'X'Y)$

It is evident that the coefficients of all the original regressors can be obtained by a linear transformation of $\hat{\beta}_z$. Furthermore, in predicting Y, it is not necessary to convert an observation's X scores into Z scores.

Finally, it should be noted that the expected value of $\hat{\beta}_x$ is only a linear transformation of the parameters of the principal components. That is,

$$\hat{\beta}_x = V_*\hat{\beta}_z = V_* [\beta_z + (Z'Z)^{-1} Z'\mu]$$
$$= V_*\beta_z + V_*(Z'Z)^{-1} Z'\mu$$

$$E(\hat{\beta}_x) = V_*\beta_z$$

The variance of $\hat{\beta}_x$ is given by

$$\begin{aligned}
\text{Var } (\hat{\beta}_x) &= E\ [\hat{\beta} - E(\hat{\beta}_x)]\ \ [\hat{\beta} - E(\hat{\beta}_x)]' \\
&= E\ [V_*(Z'Z)^{-1}\ Z'\mu\mu'Z(Z'Z)^{-1}\ V_*'] \\
&= \sigma^2 V_*\ (Z'Z)^{-1}\ V_*' \\
&= \sigma^2 V_* \Lambda_*^{-1} V_*'
\end{aligned}$$

where the variance of the disturbance σ^2 is estimated by $S^2 = \frac{(Y-\hat{Y})\ (Y-\hat{Y})'}{n-(k+p+1-m)}$